

PERCOLATION WITH SMALL CLUSTERS ON RANDOM GRAPHS

MUSTAZEE RAHMAN

ABSTRACT. We consider the problem of finding an induced subgraph in a random d -regular graph such that its components have bounded size as the size of the graph gets arbitrarily large. We show that for any threshold τ , the largest size density of such an induced subgraph with component sizes bounded by τ is at most $2(\log d)/d$ for asymptotically large d . A matching lower bound is known for independent sets.

We prove an analogous result for sparse Erdős-Rényi graphs, and provide upper and lower bounds for the largest size density on random 3-regular graphs.

1. INTRODUCTION

Consider the following problem on random d -regular graphs and Erdős-Rényi graphs of expected average degree d . A subset S of a graph G is a **percolation set with clusters of size at most τ** if all the components of the induced subgraph $G[S]$ have size at most τ . For example, independent sets have clusters of size exactly 1, and induced matchings have clusters of size exactly 2. For fixed τ and d , what is the density, $|S|/|G|$, of the largest percolation sets S with clusters of size at most τ on the aforementioned graph ensembles? We provide an asymptotically sharp upper bound to this problem for large d . Roughly speaking, for both these graph ensembles we prove that for large d and any τ the density of the largest percolation sets is at most $2\frac{\log d}{d}$ with high probability. Precise statements follow in Section 1.2. These bounds are derived from first moment estimates. First, we explain how this bound is sharp and provide some background to the problem.

Bollobás [3] showed that the density of the largest independent sets on random d -regular graphs is at most $2\frac{\log d}{d}$ as $d \rightarrow \infty$. Frieze and Łuczak followed with matching lower bounds [8, 9]. Thus, our result shows that relaxing the problem from independent sets ($\tau = 1$) to arbitrary τ provides no improvement for large d .

Bayati, Gamarnik and Tetali [2] proved that for each d the density of the largest independent sets in a random d -regular graph, or an Erdős-Rényi graph of average degree d , on n vertices converges almost surely as $n \rightarrow \infty$. Their argument can be replicated for any τ to deduce the analogous result about convergence of the density of the largest percolation sets with clusters of size at most τ .

1.1. Preliminaries and terminology. We use the well-known configuration model as the probabilistic method to sample a random d -regular graph $\mathcal{G}_{n,d}$ on n labelled vertices. Recall that $\mathcal{G}_{n,d}$ is sampled in the following manner. Each of the n vertices emit d half-edges, and we pair up these nd half-edges uniformly at random. (We tactically assume that nd is even.) These $nd/2$ pairs of half-edges can be glued into full edges to yield a random d -regular graph. There are $(nd)!! = (nd-1)(nd-3)\cdots 3\cdot 1$ such graphs.

The resulting random graph may have loops and multiple edges, that is, it is a multigraph. However, the probability that $\mathcal{G}_{n,d}$ is a simple graph is uniformly bounded away from zero at $n \rightarrow \infty$. In fact, Bender and Canfield [5] and Bollobás [4] showed that as $n \rightarrow \infty$

$$\mathbb{P}(\mathcal{G}_{n,d} \text{ is simple}) \rightarrow \exp \left\{ -\frac{(d-1)^2 + 2(d-1)}{4} \right\}.$$

Also, conditioned on $\mathcal{G}_{n,d}$ being simple its distribution is an uniform d -regular simple graph on n vertices. It follows from these observations that any sequence of events that occur with probability tending to 1 in $\mathcal{G}_{n,d}$ (as $n \rightarrow \infty$) also occurs with probability tending to 1 for an uniformly chosen simple d -regular graph.

We denote by $\text{ER}(n, p)$ a Erdős-Rényi graph on n vertices and edge inclusion probability p . Recall that every pair of vertices $\{u, v\}$ is independently included as an edge of $\text{ER}(n, p)$ with probability p . We are interested in the case $p = d/n$ for a fixed d . The resulting graph is sparse in the sense that the expected average degree is $d(1 - 1/n)$.

For a graph G and integer $\tau \geq 1$ we define

$$\alpha^\tau(G) = \max \{ |S|/|G| : S \text{ is a percolation set with clusters of size at most } \tau \}.$$

1.2. Statement of results. For large d we prove asymptotic upper bounds on $\alpha^\tau(\mathcal{G}_{n,d})$ and $\alpha^\tau(\text{ER}(n, d/n))$ that hold with high probability as $n \rightarrow \infty$.

Theorem 1.1. *For every τ and d there exists α_d^τ such that $\limsup_{d \rightarrow \infty} \frac{\alpha_d^\tau}{(\log d)/d} \leq 2$, and*

$$\mathbb{P}(\alpha^\tau(\mathcal{G}_{n,d}) \leq \alpha_d^\tau) \rightarrow 1 \quad \text{as } n \rightarrow \infty.$$

Theorem 1.2. *For every τ and d there exists $\alpha_{\text{ER}(d)}^\tau$ such that $\limsup_{d \rightarrow \infty} \frac{\alpha_{\text{ER}(d)}^\tau}{(\log d)/d} \leq 2$, and*

$$\mathbb{P}(\alpha^\tau(\text{ER}(n, d/n)) \leq \alpha_{\text{ER}(d)}^\tau) \rightarrow 1 \quad \text{as } n \rightarrow \infty.$$

The proofs of the theorems will show that α_d^τ and $\alpha_{\text{ER}(d)}^\tau$ can be chosen independently of τ so long as $\tau = o(\sqrt{\log d / \log \log d})$, although this is not optimal.

Here is another interpretation of Theorems 1.1 and 1.2. Following the argument of Bayati, Gamarnik and Tetali [2] one can show that $\alpha^\tau(\mathcal{G}_{n,d})$ and $\alpha^\tau(\text{ER}(n, d/n))$ converge

almost surely, as $n \rightarrow \infty$, to nonrandom limits $\alpha^\tau(d)$ and $\alpha^\tau(\text{ER}(d))$, respectively. Indeed, using concentration estimates, such as Azuma's inequality, it is easy to see that $\alpha^\tau(\mathcal{G}_{n,d}) - \mathbb{E}[\alpha^\tau(\mathcal{G}_{n,d})]$ and $\alpha^\tau(\text{ER}(n, d/n)) - \mathbb{E}[\alpha^\tau(\text{ER}(n, d/n))]$ converge to zero almost surely. Then the techniques from [2] can be used unchanged to show convergence of $\mathbb{E}[\alpha^\tau(\mathcal{G}_{n,d})]$ and $\mathbb{E}[\alpha^\tau(\text{ER}(n, d/n))]$. Theorems 1.1 and 1.2 are equivalent to the statement that

$$\limsup_{d \rightarrow \infty} \frac{\alpha^\tau(d)}{(\log d)/d} \leq 2 \quad \text{and} \quad \limsup_{d \rightarrow \infty} \frac{\alpha^\tau(\text{ER}(d))}{(\log d)/d} \leq 2.$$

Following the lower bound of Frieze and Łuczak [8, 9] it follows that

$$\lim_{d \rightarrow \infty} \frac{\alpha^\tau(d)}{(\log d)/d} = 2 \quad \text{and} \quad \lim_{d \rightarrow \infty} \frac{\alpha^\tau(\text{ER}(d))}{(\log d)/d} = 2.$$

For random 3-regular graphs we also provide upper and lower bounds on $\alpha(\tau, 3)$ via combinatorial arguments.

Theorem 1.3. *For any τ , $\frac{\tau-1}{\tau+1} \times 0.5491 \leq \alpha(\tau, 3) \leq 3/4$. In particular, if $\alpha(3) = \sup_\tau \alpha^\tau(3) = \lim_{\tau \rightarrow \infty} \alpha^\tau(3)$ then $0.5491 \leq \alpha(3) \leq 3/4$.*

For independent sets, McKay [12] has shown that $\alpha^1(3) \leq 0.4554$ and more recently this has been improved to $\alpha^1(3) \leq 0.4509$ in [1]. In the other direction, in [6] it is shown that $\alpha^1(3) \geq 0.4361$ and this has been improved to $\alpha^1(3) \geq 0.4375$ in [10]. To the best of our knowledge these are the best known bounds for $\alpha^\tau(3)$ for any value of τ .

We prove Theorems 1.1 and 1.3 in Section 2, and prove Theorem 1.2 in Section 3. We conclude in Section 4 with problems and comments on future directions for this work.

2. PERCOLATION ON RANDOM d -REGULAR GRAPHS

Let E denote the event that $\mathcal{G}_{n,d}$ contains a percolation set of size αn with clusters of size at most τ . We will bound the probability of E by using the first moment method.

Let X_n denote the number of loops and multiple edges in $\mathcal{G}_{n,d}$. It is a well-known fact (cf. chapter 9.2, [11]) that X_n converges in distribution to a Poisson random variable of finite mean. In particular, for any $\epsilon > 0$ there exists a constant $C_{d,\epsilon}$ such that $\mathbb{P}(X_n > C_{d,\epsilon}) \leq \epsilon$ for all n .

Let Z be the number of percolation sets of $\mathcal{G}_{n,d}$ of size αn with clusters of size at most τ . Clearly,

$$\mathbb{P}(E) \leq \mathbb{P}(E; X_n \leq C_{d,\epsilon}) + \epsilon \leq \mathbb{E}[Z; X_n \leq C_{d,\epsilon}] + \epsilon. \quad (2.1)$$

Now, if S is a percolation set with clusters of size at most τ then the number of edges in $\mathcal{G}_{n,d}[S]$ is at most $\binom{\tau}{2}|S| + X_n$. Indeed, if we ignore the loops and multiple edges of $\mathcal{G}_{n,d}$ then the subgraph $G[S]$ is a simple graph where each component contains at most $\binom{\tau}{2}$ edges.

Since there are at most $|S|$ components, there are at most $\binom{\tau}{2}|S|$ edges in $\mathcal{G}_{n,d}[S]$ excluding the loops and multiple edges. However, there are at most X_n edges of the latter kind.

Let Z_i be the number of subsets $S \subset V(\mathcal{G}_{n,d})$ such that $|S| = \alpha n$ and the number of edges in $\mathcal{G}_{n,d}[S]$ is i . The observation above implies that

$$\mathbb{E}[Z; X_n \leq C_{d,\epsilon}] \leq \sum_{i=0}^{\binom{\tau}{2}\alpha n + C_{d,\epsilon}} \mathbb{E}[Z_i]. \quad (2.2)$$

In order to calculate $\mathbb{E}[Z_i]$ we must first introduce some notation. For subsets $S, T \subset V(\mathcal{G}_{n,d})$ let

$$m(S, T) = \frac{|(u, v) : u \in S, v \in T, \{u, v\} \in E(\mathcal{G}_{n,d})|}{nd}.$$

If $\vec{e} = (e_1, e_2)$ is a uniform random directed edge of $\mathcal{G}_{n,d}$ then $m(S, T) = \mathbb{P}(e_1 \in S, e_2 \in T)$. Note that $m(S, T) = m(T, S)$, $m(S, S) = 2|E(\mathcal{G}_{n,d}[S])|/(nd)$ and if S and T are disjoint then $m(S, T)$ is the number of edges in $\mathcal{G}_{n,d}$ from S to T divided by nd . The **edge profile** of S associated to $\mathcal{G}_{n,d}$ is the 2×2 matrix

$$M(S) = \begin{bmatrix} m(S, S) & m(S, V \setminus S) \\ m(V \setminus S, S) & m(V \setminus S, V \setminus S) \end{bmatrix}$$

where $V = V(\mathcal{G}_{n,d})$. Observe that the entries of $M(S)$ sum to 1 and that the marginal distribution of $M(S)$ along the rows, or columns, is $(|S|/n, 1 - |S|/n)$. If we fix the edge profile matrix $M(\alpha, i)$ defined as

$$M(\alpha, i) = \begin{bmatrix} \frac{2i}{nd} & \alpha - \frac{2i}{nd} \\ \alpha - \frac{2i}{nd} & 1 - 2\alpha + \frac{2i}{nd} \end{bmatrix}$$

then Z_i is the number of subsets S that satisfy $M(S) = M(\alpha, i)$. Finally, the entropy of a finitely supported probability distribution π is

$$H(\pi) = \sum_x -\pi(x) \log \pi(x).$$

Lemma 2.1. *Up to a polynomial factor $\text{poly}(n, d)$ the expectation*

$$\mathbb{E}[Z_i] \leq \text{poly}(n, d) \exp \left\{ n \left[\frac{d}{2} H(M(\alpha, i)) - (d-1) H(\alpha, 1-\alpha) \right] \right\}.$$

Proof.

$$\mathbb{E}[Z_i] = \sum_{S: |S|=\alpha n} \mathbb{P}(M(S) = M(\alpha, i)) = \binom{n}{\alpha n} \mathbb{P}(M(\{1, \dots, \alpha n\}) = M(\alpha, i)). \quad (2.3)$$

The number of pairings in the configuration model satisfying $M(\{1, \dots, \alpha n\}) = M(\alpha, i)$ is

$$\binom{\alpha nd}{\alpha nd - 2i} \binom{(1-\alpha)nd}{\alpha nd - 2i} (\alpha nd - 2i)! (2i)! ((1-2\alpha)nd + 2i)!.$$

As each pairing occurs with probability $1/(nd)!!$, $\mathbb{P}(M(\{1, \dots, \alpha n\}) = M(\alpha, i))$ is

$$\binom{\alpha nd}{\alpha nd - 2i} \binom{(1-\alpha)nd}{\alpha nd - 2i} (\alpha nd - 2i)! (2i)! ((1-2\alpha)nd + 2i)! \times \frac{1}{(nd)!!}.$$

Since $m!! = \frac{m!}{2^{m/2}(m/2)!}$ we can simplify the above to

$$\frac{(\alpha nd)! ((1-\alpha)nd)! (nd/2)! 2^{\alpha nd - 2i}}{(\alpha nd - 2i)! i! \left(\frac{(1-2\alpha)}{2}nd + i\right)! (nd)!}.$$

Using Stirling's approximation of $m! \sim \sqrt{2\pi m}(m/e)^m$ we can deduce that the expression above is

$$\text{poly}(n, d) \exp \left\{ \frac{d}{2} H(M(\alpha, i)) - dH(\alpha, 1-\alpha) \right\}.$$

Also, Stirling approximation implies $\binom{n}{\alpha n} = O(\text{poly}(n)) \exp \{n H(\alpha, 1-\alpha)\}$. The statement of the lemma now follows from the expression for $\mathbb{E}[Z_i]$ in (2.3). \square

Due to its importance and prevalence in the argument we name the expression

$$\frac{d}{2} H(M(\alpha, i)) - (d-1)H(\alpha, 1-\alpha)$$

the **entropy functional**, ignoring dependencies on the variables. We conclude from Lemma 2.1 and (2.2) that in order to bound $\mathbb{E}[Z; X_n \leq C_{d,\epsilon}]$ from above it suffices to bound the entropy functional for all $0 \leq i \leq \binom{t}{2}\alpha n + C_{d,\epsilon}$. We make the tactical assumption that $\tau \leq \sqrt{d}$, which ensures that the entries of $M(\alpha, i)$ are nonnegative for large n . If $\tau > \sqrt{d}$ then we simply set $\alpha_d^\tau = 1$.

For $\tau \leq \sqrt{d}$ define the quantities $\gamma_d^\tau(\alpha)$ and α_d^τ by

$$\begin{aligned} \gamma_d^\tau(\alpha) &= \limsup_{n \rightarrow \infty} \max_{0 \leq i \leq \binom{\tau}{2}\alpha n + C_{d,\epsilon}} \frac{d}{2} H(M(\alpha, i)) - (d-1)H(\alpha, 1-\alpha) \\ \alpha_d^\tau &= \sup\{\alpha : \gamma_d(\alpha) \geq 0, 0 \leq \alpha \leq 1\}. \end{aligned}$$

From (2.2) we see that $\mathbb{E}[Z; X_n \leq C_{d,\epsilon}] \leq [\binom{\tau}{2}n + C_{d,\epsilon}] e^{n\gamma_d^\tau(\alpha)}$. Thus, if $\gamma_d^\tau(\alpha) < 0$ then from (2.1) we deduce that $\lim_{n \rightarrow \infty} \mathbb{P}(E) = 0$. In the rest of the section we will bound $\gamma_d^\tau(\alpha)$ to show that $\alpha_d^\tau \leq (2 + o(1)) \frac{\log d}{d}$ in the sense that $\limsup_{d \rightarrow \infty} \frac{\alpha_d^\tau}{(\log d)/d} \leq 2$.

First, we show that in order for $\gamma_d^\tau(\alpha) \geq 0$ we must have $\alpha_d^\tau \rightarrow 0$ as $d \rightarrow \infty$. Suppose otherwise, so that $\limsup_{d \rightarrow \infty} \alpha_d^\tau = \alpha_\infty > 0$. For a matrix M let $\|M\|$ denote its largest entry in absolute value. Then

$$\limsup_{d \rightarrow \infty} \limsup_{n \rightarrow \infty} \max_{0 \leq i \leq \binom{\tau}{2}\alpha n + C_{d,\epsilon}} \|M(\alpha_d, i) - M(\alpha_\infty, 0)\| = 0.$$

By continuity of the entropy function H it follows that

$$\limsup_{d \rightarrow \infty} \gamma_d^\tau(\alpha_d^\tau)/d \leq \frac{1}{2} H(M(\alpha_\infty, 0)) - H(\alpha_\infty, 1-\alpha_\infty).$$

However, $(1/2)H(M(x, 0)) - H(x, 1-x) = (1-x)\log(1-x) - (1/2)(1-2x)\log(1-2x)$, and this is negative for $x > 0$. This can be seen by noting that the derivative of the expression is negative for $x > 0$ and the expression vanishes at $x = 0$.

Having concluded that $\alpha_d^\tau \rightarrow 0$ we now bound the entropy functional and analyze its asymptotic behaviour for large d . Observe that for any $0 \leq i \leq \binom{\tau}{2}\alpha n + C_{d,\epsilon}$, there exists a $0 \leq t \leq \tau^2$ such that

$$\lim_{n \rightarrow \infty} M(\alpha, i) = \begin{bmatrix} \alpha \frac{t}{d} & \alpha - \alpha \frac{t}{d} \\ \alpha - \alpha \frac{t}{d} & 1 - 2\alpha + \alpha \frac{t}{d} \end{bmatrix} =: \mathbf{M}(\alpha, t).$$

From the continuity of the entropy H we conclude that

$$\gamma_d^\tau(\alpha) \leq \sup_{0 \leq t \leq \tau^2} \frac{d}{2} H(\mathbf{M}(\alpha, t)) - (d-1)H(\alpha, 1-\alpha).$$

Set $h(x) = -x\log(x)$ for $0 \leq x \leq 1$. To analyze the entropy functional for large d we are going to use the following two properties of $h(x)$. First, $h(xy) = xh(y) + yh(x)$ and second, by Taylor expansion, $h(1-x) = x - x^2/2 + O(x^3)$ as $x \rightarrow 0$. From this we conclude the following.

$$\begin{aligned} H(M(\alpha, t)) &= h(\alpha \frac{t}{d}) + 2h(\alpha - \alpha \frac{t}{d}) + h(1 - 2\alpha + \alpha \frac{t}{d}) \\ &= 2[h(\alpha) + \alpha - \alpha^2] + \frac{t}{d}[\alpha - h(\alpha) + \alpha \log d + 2\alpha^2] + O(\alpha^3 + \frac{t^3}{d^3}), \end{aligned} \quad (2.4)$$

$$H(\alpha, 1-\alpha) = h(\alpha) + \alpha - \frac{1}{2}\alpha^2 + O(\alpha^3). \quad (2.5)$$

From (2.4) and (2.5) we conclude that

$$\frac{d}{2}H(\mathbf{M}(\alpha, t)) - (d-1)H(\alpha, 1-\alpha) = h(\alpha) - \frac{d}{2}\alpha^2 + \frac{t}{2}[\alpha \log d - h(\alpha)] + O(t\alpha + d\alpha^3 + \frac{t^3}{d^2}). \quad (2.6)$$

In order to understand the asymptotic behaviour of (2.6) α needs to be of order $\frac{\log d}{d}$. Set $\alpha = \beta \frac{\log d}{d}$; we omit expressing the dependence of β on d . The right hand side of (2.6) becomes

$$\left(\beta - \frac{1}{2}\beta^2\right) \frac{\log^2 d}{d} + O\left(\frac{t\beta^2 \log d}{d} + \frac{t\beta \log d \log \log d}{d} + \frac{\beta^3 \log^3 d}{d^2} + \frac{t^3}{d^2}\right).$$

Taking $\tau = o(\sqrt{\log d / \log \log d})$ as $d \rightarrow \infty$ and noting that $t \leq \tau^2$ we see that the above it at most

$$\beta \frac{\log^2 d}{d} \left[1 - \frac{\beta}{2} + o(1) + o\left(\frac{\beta}{\log \log d}\right) + O\left(\frac{\beta^2 \log d}{d}\right)\right].$$

Recall that $\alpha \rightarrow 0$ with d and so $\beta = o(d/\log d)$. Therefore, $o(\beta/\log \log d) + O(\beta^2 \log d/d) = o(\beta)$ as $d \rightarrow \infty$. With this observation and the bound above we conclude that the entropy

functional satisfies

$$\sup_{0 \leq t \leq \tau^2} \frac{d}{2} H(\mathbf{M}(\beta \frac{\log d}{d}, t)) - (d-1) H(\beta \frac{\log d}{d}, 1 - \beta \frac{\log d}{d}) = (\beta - \frac{1}{2}\beta^2) \frac{\log^2 d}{d} + o(\frac{\beta^2 \log^2 d}{d}).$$

Consequently, $\gamma_d^\tau(\beta \frac{\log d}{d}) < 0$ for all large d unless $\beta \leq 2 + o(1)$. Thus, $\alpha_d^\tau \leq (2 + o(1)) \frac{\log d}{d}$. This establishes Theorem 1.1.

2.1. Bounds for random 3-regular graphs. We establish the bounds given in Theorem 1.3. The upper bound is a consequence for the following useful observation about boundaries of induced trees in a 3-regular graph.

Lemma 2.2. *Let T be a finite, connected, induced subgraph of a 3-regular graph G . Define the boundary of T as $\partial T = \{v \in V(G) \setminus V(T) : \exists u \in V(T) \text{ with } \{v, u\} \in E(G)\}$. If the subgraph induced by $T \cup \partial T$ contains no cycles then $|\partial T| = |T| + 2$.*

Proof. We can proceed by induction on $|T|$. Note that T is a tree by assumption and so it contains a leaf vertex v . Therefore, v has two neighbours in ∂T and these two vertices are not incident to any vertices of T other than v . The latter property holds because $T \cup \partial T$ contains no cycles.

Let $T' = T \setminus \{v\}$, which satisfies the same hypothesis as T . Observe that $|\partial T'| = |\partial T| - 1$. Indeed, removing v from T adds v to $\partial T'$ and removes its two neighbours, which are in ∂T , from $\partial T'$. The remainder of ∂T is retained in $\partial T'$. Therefore, $|\partial T| - |T| = |\partial T'| - |T'|$. It follows from induction that if w is any vertex in G then $|\partial T| - |T| = |\partial\{w\}| - 1 = 2$. \square

Proof of the upper bound in Theorem 1.3. Now suppose a 3-regular graph G has girth at least $3\tau + 1$, that is, the length of the shortest cycle in G is at least $3\tau + 1$. If S is a percolation set of G with clusters of size at most τ then the components of S satisfy the conditions of Lemma 2.2. Indeed, any component \mathcal{C} can have at most $3|\mathcal{C}| \leq 3\tau$ vertices in $\mathcal{C} \cup \partial\mathcal{C}$ and hence no cycles. Expressing S as a disjoint union of components we deduce from Lemma 2.2 that

$$|S| < \sum_{\text{components } \mathcal{C}} |\partial\mathcal{C}|.$$

On the other hand, only the vertices in $V(G) \setminus S$ contribute to the sum above and every such vertex is counted at most thrice. Consequently, $|S| < 3(|G| - |S|)$, or equivalently, $|S| < (3/4)|G|$.

So far the argument shows that $\alpha^\tau(G) < 3/4$ if G has girth at least $3\tau + 1$. It is well-known that the number of cycles of length at most ℓ in $\mathcal{G}_{n,3}$, say $X_n(\ell)$, satisfies the property that $X_n(\ell)/n \rightarrow 0$ almost surely as $n \rightarrow \infty$ (cf. chapter 9.2, [11]). If we remove a vertex of $\mathcal{G}_{n,3}$ from each cycle of length at most 3τ and carry out the analysis above then it is easy to see that $\alpha^\tau(\mathcal{G}_{n,3}) \leq 3/4 + O(X_n(3\tau + 1)/n)$. Thus, $\alpha^\tau(3) \leq 3/4$ for all τ .

Proof of the lower bound in Theorem 1.3. The lower bound uses the following observation. Let T be a induced subgraph of a 3-regular graph G such that every vertex in T has at least one neighbour in $V(G) \setminus T$. Then the components of T are cycles or paths because every vertex in T has degree at most 2 in T .

We can find such a set T by taking the complement of a maximal independent set, for example. Within such T we can find a percolation set S with clusters of size at most τ and $|S| \geq \frac{\tau-1}{\tau+1}|T|$. To do this consider any component \mathcal{C} of T of size $\ell = (\tau+1)q + r$ with $0 \leq r \leq \tau$. Since \mathcal{C} is either a cycle or a path we can decompose it into q disjoint, consecutive paths of length $\tau+1$ each along with a final path of length r . We pick the initial segment of length τ from each of the paths of length $\tau+1$ and the initial segment of length $r-1$ (provided $r > 0$) from the last path. The union of these segments is a percolation set with clusters of size at most τ . A simple calculation shows that the size of the union is at least $\frac{\tau-1}{\tau+1}\ell$. Taking the union of these percolation sets over all components of T gives us the desired set S .

It is shown in [1] that maximal independent sets of $\mathcal{G}_{n,3}$ have size at most $0.4509n$ with probability tending to 1 as $n \rightarrow \infty$, that is, $\alpha^1(3) \leq 0.4509$. Taking T to be any maximal independent set in $\mathcal{G}_{n,3}$ we conclude from the above that $\alpha^\tau(\mathcal{G}_{n,3}) \geq \frac{\tau-1}{\tau+1} \times 0.5491$ with probability tending to 1, and thus, $\alpha^\tau(3) \geq \frac{\tau-1}{\tau+1} \times 0.5491$.

3. PERCOLATION ON ERDŐS-RÉNYI GRAPHS

We will prove Theorem 1.2. Let $G \sim \text{ER}(n, d/n)$. For $\tau \geq 1$ let $S \subset V(G)$ be a percolation set with clusters of size at most τ such that $|S| = \alpha n$. Let Z be the number of such S and for $\alpha_1, \dots, \alpha_\tau$, let $Z(\alpha_1, \dots, \alpha_\tau)$ denote the number of such S with the property that $G[S]$ contains $\alpha_k n$ components of size k for each $1 \leq k \leq \tau$. Note that $\alpha = \sum_k k \alpha_k$. Clearly,

$$Z = \sum_{\alpha_1, \dots, \alpha_\tau} Z(\alpha_1, \dots, \alpha_\tau)$$

where the sum is over all admissible α_k .

As in the case of d -regular graphs we will estimate $\mathbb{E}[Z]$ and use the first moment method to bound $\mathbb{P}(Z \geq 1)$. The expectation is of exponential order and we are thus interested in $(\log \mathbb{E}[Z])/n$. There are at most n^τ admissible values of $\alpha_1, \dots, \alpha_\tau$ because each $\alpha_k n$ is an integer between 1 and n . Therefore,

$$\frac{\log \mathbb{E}[Z]}{n} \leq \max_{\alpha_1, \dots, \alpha_\tau \text{ admissible}} \frac{\log \mathbb{E}[Z(\alpha_1, \dots, \alpha_\tau)]}{n} + O\left(\frac{\log n}{n}\right). \quad (3.1)$$

We define α_d^{ER} to be the largest $\alpha \in [0, 1]$ such that the limit supremum in n of the right hand side of (3.1) is non-negative. We will show that $\alpha_d^{\text{ER}} \leq (2 + o(1))\frac{\log d}{d}$ as $d \rightarrow \infty$, which establish Theorem 1.2.

We now bound $\mathbb{E}[Z(\alpha_1, \dots, \alpha_\tau)]$. This expectation is a sum of probabilities of configurations where each configuration stipulates a unique way a percolation set S with clusters of size at most t can arise and satisfy all the required constraints. In the following we describe these configurations.

First, there exists a partition of $V(G)$ into ordered cells $P_1, \dots, P_\tau, P_{\tau+1}$ such that $|P_k| = k\alpha_k n$ for $1 \leq k \leq \tau$. The cell P_k consists of vertices that will form the components of size k . Then each P_k is further partitioned into $\alpha_k n$ unordered cells of size k . These mini-cells will form the components of size k . Therefore, the mini-cells must be connected in G and there can not exist any edge in G that connects two different mini-cells.

All the configurations have the same probability of begin present in $\text{ER}(n, d/n)$, and thus, $\mathbb{E}[Z(\alpha_1, \dots, \alpha_\tau)]$ is the product of the number of configurations and the probability that a configuration is present. We calculate these two quantities.

Number of configurations. The number of partitions of the vertex set $V(G)$ into ordered cells $P_1, \dots, P_{\tau+1}$ is the multinomial term

$$\binom{n}{k\alpha_k n; 1 \leq k \leq \tau, (1-\alpha)n}. \quad (3.2)$$

Recall that $\alpha = \sum_k k\alpha_k$.

Given such a partition, the number of partitions of P_k into $\alpha_k n$ unordered mini-cells of size k is $\frac{(k\alpha_k n)!}{(k!)^{\alpha_k n} (\alpha_k n)!}$. Division by $(\alpha_k n)!$ ensures that the mini-cells are unordered. The total number of unordered partitions of the mini-cells is thus

$$\prod_{k=1}^{\tau} \frac{(k\alpha_k n)!}{(k!)^{\alpha_k n} (\alpha_k n)!}. \quad (3.3)$$

The total number of configurations is the product of (3.2) and (3.3).

Probability of a configuration. The probability that a mini-cell of size k is connected is the probability that $\text{ER}(k, d/n)$ is connected. Let $p_k = \mathbb{P}(\text{ER}(k, d/n) \text{ is connected})$. There are $\alpha_k n$ occurrences of p_k and their contribution to the probability of a configuration is $p_k^{\alpha_k n}$. The probability that all the mini-cells are connected is

$$\prod_{k=1}^{\tau} p_k^{\alpha_k n}. \quad (3.4)$$

That the mini-cells remain disconnected in the induced subgraph is ensured by requiring that all edges that could connect vertices from different mini-cells are not present in G . By elementary counting the number of such edges is $[(\alpha n)^2 - \sum_k k^2 \alpha_k n]/2$. Thus, the probability associated to this event is

$$\left(1 - \frac{d}{n}\right)^{\frac{(\alpha n)^2 - \sum_k k^2 \alpha_k n}{2}}. \quad (3.5)$$

The probability of any configuration being present is the product of (3.4) and (3.5). Hence, $\mathbb{E}[Z(\alpha_1, \dots, \alpha_\tau)]$ equals

$$\binom{n}{k\alpha_k n; 1 \leq k \leq \tau, (1-\alpha)n} \prod_{k=1}^{\tau} \frac{(k\alpha_k n)!}{(k!)^{\alpha_k n} (\alpha_k n)!} \prod_{k=1}^{\tau} p_k^{\alpha_k n} \left(1 - \frac{d}{n}\right)^{\frac{(\alpha n)^2 - \sum_k k^2 \alpha_k n}{2}}.$$

We will estimate all of the terms above up to exponential order. For large d the main contribution to the expectation comes from (3.2) and (3.5). Using Stirling's approximation it is easy to see that up to errors of order $O(\log n/n)$ the multinomial term in (3.2) is of exponential order

$$H(\alpha_i, 2\alpha_2, \dots, k\alpha_k, 1-\alpha) = \sum_{k=1}^{\tau} -k\alpha_k \log(k\alpha_k) - (1-\alpha) \log(1-\alpha).$$

Since $1 - \frac{d}{n} \leq e^{-d/n}$ we see that the probability in (3.5) is at most of exponential order

$$-\frac{d\alpha^2}{2} + O(1/n).$$

Now we account for the remaining terms. Observe that p_k is the probability that at least one of the k^{k-2} trees on k vertices appear in $\text{ER}(k, d/n)$. The probability that any given tree appears is $(d/n)^{k-1}$. By a union bound we conclude that $p_k \leq k^{k-2}(d/n)^{k-1}$. Therefore, up to exponential order, the term in (3.4) is at most

$$\sum_{k=1}^{\tau} \alpha_k [(k-1) \log\left(\frac{d}{n}\right) + (k-2) \log k].$$

Finally the term in (3.3) can be estimated by Stirling's approximation. Its exponential order is

$$\sum_{k=1}^{\tau} \alpha_k [(k-1) \log\left(\frac{\alpha_k n}{e}\right) + k \log k - \log k!] + O\left(\frac{1}{n}\right).$$

We conclude that the contribution of (3.3) and (3.4) to the expectation is at most of exponential order

$$\sum_{k=1}^{\tau} \alpha_k [(k-1) \log\left(\frac{\alpha_k d}{e}\right) + (2k-2) \log k - \log k!] + O\left(\frac{1}{n}\right).$$

Combining the above with the contribution from (3.2) and (3.5) and using (3.1) we see that

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{\log \mathbb{E}[Z]}{n} &\leq \sup_{\substack{\alpha_1, \dots, \alpha_\tau: \\ \alpha_1 + 2\alpha_2 + \dots + \tau\alpha_\tau = \alpha}} \sum_{k=1}^{\tau} -k\alpha_k \log(k\alpha_k) - (1-\alpha) \log(1-\alpha) - \frac{d\alpha^2}{2} \quad (3.6) \\ &\quad + \sum_{k=1}^{\tau} \alpha_k [(k-1) \log\left(\frac{\alpha_k d}{e}\right) + (2k-2) \log k - \log k!]. \end{aligned}$$

Similar to the case for d -regular graphs, the right hand side of (3.6) is negative for all large d unless $\alpha \rightarrow 0$ with d . Indeed, if α is bounded away from zero then the dominating term is $-(d/2)\alpha^2$. To analyze (3.6) for large d write $\alpha_k = \hat{\alpha}_k \frac{\log d}{d}$ and $\alpha = \hat{\alpha} \frac{\log d}{d}$. One can check that $-k\alpha_k \log(k\alpha_k) = k\hat{\alpha}_k \frac{\log^2 d}{d} + O(\frac{\log d}{d})$. Similarly, from Taylor expansion $-(1-\alpha) \log(1-\alpha) = O(\hat{\alpha} \frac{\log d}{d})$. This shows that

$$\sum_{k=1}^{\tau} -k\alpha_k \log(k\alpha_k) - (1-\alpha) \log(1-\alpha) - \frac{d\alpha^2}{2} = (\hat{\alpha} - \frac{1}{2}\hat{\alpha}^2) \frac{\log^2 d}{d} + O(\hat{\alpha} \tau \frac{\log d}{d}).$$

On the other hand it is easy to check that the contribution from the remaining term in (3.6) is $O(\hat{\alpha} \log \hat{\alpha} \log \tau \log d \log \log d/d)$. Consequently,

$$\limsup_{n \rightarrow \infty} \frac{\log \mathbb{E}[Z]}{n} = (\hat{\alpha} - \frac{1}{2}\hat{\alpha}^2) \frac{\log^2 d}{d} + O\left(\frac{(\hat{\alpha} \log \hat{\alpha}) \tau \log d \log \log d}{d}\right).$$

We conclude from this that unless $\hat{\alpha} \leq 2 + o(1)$ as $d \rightarrow \infty$ then $\limsup_{n \rightarrow \infty} (\log \mathbb{E}[Z])/n < 0$ for all large d . As a result, $\alpha_d^{\text{ER}} \leq (2 + o(1)) \frac{\log d}{d}$ and Theorem 1.2 follows.

4. CONCLUDING REMARKS

It is natural to consider percolation sets in $\mathcal{G}_{n,d}$ or $\text{ER}(n, d/n)$ with *finite clusters* as $n \rightarrow \infty$ in the sense of the following limiting quantities:

$$\alpha(d) = \sup_{\tau} \alpha^{\tau}(d) = \lim_{\tau \rightarrow \infty} \alpha^{\tau}(d) \text{ and } \alpha(\text{ER}(d)) = \sup_{\tau} \alpha^{\tau}(\text{ER}(d)) = \lim_{\tau \rightarrow \infty} \alpha^{\tau}(\text{ER}(d)).$$

Our first moment estimates are not strong enough to show that $\alpha(d)$ or $\alpha(\text{ER}(d))$ are asymptotically at most $(2 + o(1)) \frac{\log d}{d}$. However, we expect this to be the case. Since $\mathcal{G}_{n,d}$ and $\text{ER}(n, d/n)$ have a small number of cycles of fixed lengths, and as trees contain independent sets of density at least $1/2$, one can show that $\alpha(d) \leq 2\alpha^1(d)$ and $\alpha(\text{ER}(d)) \leq 2\alpha^1(\text{ER}(d))$. Hence, a reasonable approach would be to show that for every d one has the bound $\alpha(d) = (1 + o(1)) \alpha^{\tau}(d)$ as $\tau \rightarrow \infty$, and analogously for $\alpha(\text{ER}(d))$.

It would also be of interest to obtain sharp estimates for $\alpha(d)$ for both small and sufficiently large, but fixed, values of d . Some lower bounds for $\alpha(3)$ are given in [6] and the approach provides good lower bounds for small values of d as well. On the other hand, [7] expresses the independence ratio $\alpha^1(d)$ implicitly, but with exact formulae, for all sufficiently large values of d . Their approach is guided by predictions from statistical physics about the structure of independent sets in $\mathcal{G}_{n,d}$ and these ideas may apply to percolation sets as well.

ACKNOWLEDGEMENTS

The author is grateful to Viktor Harangi for the argument leading to the upper bound in Theorem 1.3, in particular, for Lemma 2.2. The author also thanks Balint Virág for suggesting the problem. The author's research is supported by a NSERC CGS grant.

REFERENCES

- [1] J. Barbier, F. Krzakala, L. Zdeborova, and P. Zhang The hard-core model on random graphs revisited, *J. Phys.: Conf. Ser.*, 473 012021 (2013), [arXiv:1306.4121](#).
- [2] M. Bayati, D. Gamarnik, and P. Tetali, Combinatorial approach to the interpolation method and scaling limits in sparse random graphs, *Annals of Probability*, 41 (2013) pp. 4080–4115, [arXiv:0912.2444](#).
- [3] B. Bollobás, The independence ratio of regular graphs, *Proc. Amer. Math. Soc.*, 83(2) (1981) pp. 433–436.
- [4] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled regular graphs, *Eur. J. Combin.*, 1 (1980) pp. 311–316.
- [5] E. A. Bender and E. R. Canfield, The asymptotic number of labelled graphs with given degree sequences, *J. Combin. Theory, Ser. A* 24 (1978) pp. 296–307.
- [6] E. Csóka, B. Gerencsér, V. Harangi, and B. Virág, Invariant Gaussian processes and independent sets on regular graphs of large girth, *preprint* (2013), [arXiv:1305.3977](#).
- [7] J. Ding, A. Sly, N. Sun, Maximum independent sets on random regular graphs, *preprint* (2013), [arXiv:1310.4787](#).
- [8] A.M. Frieze, On the independence number of random graphs, *Discrete Mathematics*, 81 (1990) pp. 171–175.
- [9] A.M. Frieze and T. Łuczak, On the independence and chromatic numbers of random regular graphs, *J. Combin. Theory, Ser. B* 54 (1992) pp. 123–132.
- [10] C. Hoppen and N. Worlmal, Local algorithms, regular graphs of large girth, and random regular graphs, *preprint* (2013), [arXiv:1308.0266](#).
- [11] S. Janson, T. Łuczak, and A. Rucinski, Random graphs, John Wiley and Sons, Inc., 2000.
- [12] B.D. McKay, Independent sets in regular graphs of high girth, *Ars Combin.*, 23A (1987) pp. 179–185.

(Mustazee Rahman) DEPARTMENT OF MATHEMATICS, UNIVERSITY OF TORONTO, 40 ST. GEORGE STREET, TORONTO, ON M5S 2E4, CANADA

E-mail address, Mustazee Rahman: mustazee@math.toronto.edu